基于微博数据探究宗族文化观念与单人家庭户地域差异的关系

周子力[1, 2]　　李泓仪[1, 2]　　朱廷劭[1, 2]
[1]（中国科学院心理研究所 北京 100101）
[2]（中国科学院大学心理学系 北京 100049）

摘要：近年来，单人家庭户比例快速上升，折射出中国正在经历的家庭和居住模式变迁过程，这很可能受到文化因素，尤其是中国传统宗族文化的深刻影响。本研究尝试引入基于微博文本数据的词典统计方法，结合心理语言学原理测量用户发表内容中呈现的宗族观念水平，并综合其他变量探究宗族观念水平在单人家庭户地域差异中的作用。通过检验，宗族观念词典测量结果与人工评分具有相当的关联性，满足初步用于微博文本数据分析的要求。在横断面回归分析中发现，同时纳入宗族观念及其与地级市人口规模、人均 GDP 的交互作用时，回归模型有显著提升，且宗族观念水平与单人户比例负向关联。本研究的成果提示了宗族文化在单人户地域差异形成中的作用，为弥合家庭模式变迁理论与实证差距提供了新的思路，并为未来研究中定量化使用宗族文化指标开辟了新的路径。

关键词：宗族文化　单人家庭户　词典　社交媒体

1　引言

家庭户是指以家庭成员关系为主、居住一处共同生活的人组成的户，是中国人口统计中代表独立家庭实际居住生活模式的基本统计指标(中华人民共和国国家统计局, 2011)。伴随着中国的高速发展和深刻的社会变化，居住模式和家庭户结构也发生了深刻演变，根据第七次全国人口普查（2020 年）数据，单人家庭户占比达到 25.39%，已成为中国居民重要的居住或生活方式(王跃生, 2023)。单人户家庭通常与单人居住模式紧密联系，意味着传统家庭功能弱化，单人家庭户作为一种非完整家庭结构，伴随着脆弱性和风险性问题（穆光宗，2021）。因此居住模式，尤其是显著的单人家庭户比例将直接关联到社会的稳定、人口结构的调整以及社会保障政策的制定。

中国单人家庭户从 20 世纪 80 年代至今经历了一个数量比例上快速增长，结构迅速变化的阶段。Cheung & Yeung（2015）发现 1982 至 2005 年，单身和已婚人士形成的单人户开始占据多数。自 20 世纪 80 年代以来，中国的家庭户平均规模在明显缩小，单人家庭户的比例显著增加。单人户的增长带来了一系列的社会问题，导致家庭功能减弱、代际支持能力下降以及养老问题加剧（王广州 & 周玉娇, 2021）。此外，单人居住模式对人口的影响还体现在减少了婚育队伍的人口规模，且具有快速增长的特征(王跃生, 2023)。

现有对于单人家庭户变迁的研究较多使用人口普查、专项调查和历史史料等调查资料，如历次人口普查资料、各地地方志所记载的户籍人口资料和由研究者形成的专题研究。比较典型的成果包括家庭户结构的数量变化趋势，如胡湛 & 彭希哲（2014）利用 1982-2010 年人口普查数据分析得出对于中国总体而言，家庭户数增长趋势快于人口增长趋势，家庭规模持续缩减，单人户的快速增长与老龄化、空巢化紧密关联。周长洪(2013)详细考察单人户在不同年龄段、性别和城

乡背景下的差异，得出中青年男性和老年（65 岁以上）女性更容易成为单人户，且城镇相对于农村、老年相对于中青年单人户比例更大的结论。而关于单人户空间分布特征和地域差异因素的分析则相对较少，从理论上看，家庭现代化理论被广泛采用，其认为技术发展改变了社会的生产和经济基础，导致家庭功能和个人价值体系的巨大变化，从而引发了住户结构的变化（Goode 1963；Nimkoff 1961；Parsons 1949）。此外，人口特征变化对住户结构有着直接影响（Jiang 和 O'Neill , 2007），婚姻挤压和老龄化显著地提高了单人家庭户的比例(王跃生, 2023；王广州 & 周玉娇, 2021)；制度变化对中国人口与家庭也有着尤为显著的影响，特别是生育调控政策对生育率和年龄结构起到的作用（Wang， 2011）和户籍制度与经济改革对人口迁移、家庭成员分离的作用（Liang 和 White 1996；Wen 和 Hanley 2015）。王跃生（2012）从大型直系家庭转变为小型核心家庭和单人家庭的角度出发，对中国范围省、市两级行政单位作为比较单位，得出人均收入高、外来人口比例大和老龄化水平高的地区有着更多的夫妻二人家庭和单人户家庭。然而家庭现代化解释也面临一些挑战，如曾毅等(1992)发现社会经济发展水平与单人户比例有一定相关性，但在东北地区三个省份中不成立；马春华等（2011）对五个城市的家庭研究显示，家庭发展受到非经济因素如宗族观念的关键作用，这一点与沈崇麟（2009)所认为的家庭发展水平与经济发展水平有相应梯度关系的结论相矛盾。单人户同时是家庭代际结构（一代家庭的特殊形式）和家庭类型结构（直系家庭分裂的结果）中的重要特征，而目前存在区域比较视野下的单人户研究积累相对薄弱的问题(王跃生, 2024)，因此在上述研究主要涉及经济发展水平、城乡差异和人口迁移因素等经济地理因素基础上，进一步综合文化变量发挥的重要作用。

文化影响也是中国家庭研究中常被提及的机制，其基本理论认为文化起源在决定家庭户模式方面起到了重要的作用。家庭、家族和宗族在是中国在历史上发挥极其重要作用的社会结构单位，单人家庭户变化属于家庭户演变的一个重要方面。王跃生（2016）系统地阐述了中国家庭和家庭户研究中存在的三个重要概念："家"、"家庭"、 "家庭户"，提示综合地考察"家"或"家庭"意义上的生活方式和变迁模式非常必要。"家本位"的文化传统是中国古代社会变迁中始终延续的基础性要素。在家庭结构的维系和变动过程中，分家、生育和迁移流动被认为是产生影响的重要原因，传统宗法家族制度失去法律支持后，中国家庭很快完成了核心化的过程（王跃生，2008）。分家行为是中国家庭户产生和变化的直接来源，龚为纲（2012）提出"父权"减弱，妇女地位上升和兄弟关系变化是分家模式演变的关键影响因素。因此，单人户作为传统直系家庭向更现代的家庭形式过渡的突出结果，必须考虑宗族文化因素的重要作用。而在现有研究中，尽管有研究注意到某些地区的宗族文化传统对家庭户结构变化可能有重要作用（马春华等，2011），但很少将宗族文化差异纳入地域差异和纵向变迁分析中。由于中国家庭具有受文化传统影响的突出特点，本研究希望发展宗族文化观念变量的定量化方法，并将其纳入单人户区域比较分析中，以弥合现有研究中的矛盾和不足，更全面刻画与解释中国家庭户演变发展路径。

宗族观念是个体对宗族关系、制度在思想上和行动上的认同(陈丽, 2007)。虽然经历了 20 世纪的民主主义革命和土地改革，破坏了宗法制度在土地经济模式上的生存基础，但宗法制度以及宗族文化的影响不仅在历史上广泛存在，而且在现代社会中依然在不同程度上影响着宗族成员的行为和价值观(吴祖鲲 ，王慧姝, 2014) (Peng, 2004)。中国的宗族传统形成了一套父权制社会规范，规范通过定义

家庭成员的等级顺序、确定他们的义务和责任、控制生育行为、建立生活安排以及决定家庭财产如何分配等方面，深入影响着住户结构（Wang，2008）。

宗族文化及其对家庭变迁的作用没有在学界形成稳定的共识，也缺乏广泛检验的调查数据加以佐证。对宗族文化传统的测量缺少公认的标准方法，一般采取不同的历史数据来衡量宗族发展水平，如姓氏集中度、宗祠数量等，或者以较为笼统的“文化差异”来侧面反映文化背景因素的作用，如王晓慧等(2012)对分代家庭户地域差异的分析从宗教信仰与文化传统等多重因素，Li 等（2020）采用方言区和少数民族分布地区差异来解释不同文化背景导致的住户结构类型变迁模式差异。这些变量很大程度上反映了中国宗族活动在经过历史变迁后表现出的差异性，但也存在缺少定量工具交叉佐证的情况。另一方面，由于历史数据依赖于调查的实际发生时间，并用多次调查数据时可能发生滞后问题。如中国的综合社会调查、经济普查和人口普查分别于不同年份进行。对于受到人口迁移、城镇化快速变化重大影响的人口结构问题，很可能因数据采集时间差异造成较大误差。此外，这些数据没有直接测量人群的心理特征，即宗族发展的历史水平是否与个体宗族观念水平存在必然联系，并没有受到有效的验证。为此，本研究试图尝试引入用户生成内容（UGC）分析方法，从民众行为数据获取宗族观念水平，以回答如下的研究问题：宗族文化背景如何影响单人家庭户的地域分布特征？宗族文化背景与其他经济地理变量是否存在以及存在怎样的交互作用？

新浪微博在 2009 年上线后成为中国大陆最被广泛使用的社交媒体，从微博爬取的用户数据具有时间精度高，空间覆盖范围广和生态效度高的特点，是现有研究从大数据分析入手揭示用户外显行为和内在心理机制关系的重要参考样本(Qiu et al., 2012; 苏悦等, 2021)。在自然语言处理（NLP）中，词典方法是一种基于已有的词汇资源来分析文本的技术。这种方法通常涉及到使用预定义的词汇表或词典，利用词频统计来代表具有特定意义的某一项目的相对数值，被广泛应用于情感分析、主题建模和语义解析等领域中。

因此，本研究设计分两阶段进行，研究一将使用网络社交媒体文本数据和心理语言学方法，基于微博数据建立宗族观念词典，并利用验证性实验设计考察处理方法在语料库上的效度与信度。基于此我们提出假设：H1-词水平统计结果与人类专家对文本宗族观念水平的判断具有一致性。

在此基础上，研究 2 将开展与历史资料数据进行时空精确匹配的关联分析研究，结合第六次全国人口普查数据和统计年鉴的公开宏观数据，利用横断面回归分析，综合宗族观念水平的地域差异，以及宗族观念对家庭户数量结构的可能影响。一方面我们希望考察宗族文化观念对单人户的影响，同时我们还关注宗族文化观念变量的纳入对原有理论的作用，为此我们提出两条假设：H2-宗族文化观念对单人户比例有显著的预测作用；H3-宗族文化观念与经济地理变量有显著的交互作用。

2 研究一：通过微博数据进行宗族文化观念测量

2.1 数据来源

本研究主要使用了利用 python 爬虫程序从新浪微博获取的文本数据。微博数据作为网络社交媒体，具有时间精度高，空间覆盖范围广和生态效度高的特点，是现有研究从大数据分析入手揭示用户外显行为和内在心理机制关系的重要参考样本(Qiu et al., 2012; 苏悦等, 2021)。微博样本涵盖从 2010 年到 2022 年的中

国 31 个省（包含直辖市、自治区，不含港澳台）的 25 亿条文本。

此外还参考了中国国家统计局发布的中国统计年鉴(2011)和第六次人口普查数据(2010)以及一项面向中国 17 个省 131 个地级行政区的问卷调查（样本数=6700 人）。本研究使用了问卷中四个量表题结果作为“宗族活动”的代表性指标，其分别为“D1.您的家庭所在的村庄（或您的父母所在的村庄）是否有祠堂（宗祠）？”“D2.您的家庭是否有家谱？”“D3.您的家庭所在的村庄（或您的父母所在的村庄）是否有族谱？”“D3.您的家庭所在的村庄（或您的父母所在的村庄）是否有族谱？”“D4.您一年中祭拜祖先和参与上坟的次数为？”。这四个问题的测量与编码方式见附表 1。

2.2 数据处理与统计分析

本研究主要利用经过验证的词频统计程序对文本数据进行词频统计，经过分词后对各维度包含的词进行词频统计，每一文本输入对应输出一个词频向量。

数据获取和处理后，利用 Microsoft Excel 和 IBM SPSS 27 软件包进行相关分析、一致性检验以及其他统计分析。

2.3 方法与程序

（1）初始种子词表的构建与扩充

为确定宗族文化观念文本测量的核心，本研究从核心文献出发，搜集和建立宗族文化观念词典的核心词表，主要参考了常建华(1999)对中国本土宗族研究的综述和中国农村调查(总第 2 卷·村庄类第 1 卷·华南区域第 1 卷)(徐勇 & 邓大才, 2017)。经过人工总结，主要确定了符合定义的外显标志词语作为词表的第一部分。通过核对和清洗，第一部分共包含 117 个词。

同时，由于血缘纽带和姓氏继承（冠姓权）在宗族文化中的核心地位，本研究利用组合生成的方式，向外显标志词中纳入了表征姓氏继承和近亲属的常用词。这一部分包含 332 个词。基于核心关键词表，利用人工扩充和大语言模型扩充相结合的方法，根据词义进行同义词、近义词扩充，共生成了 1246 个词作为扩充词表。

（2）基于相关指标的词语筛选

根据 Gong 等人(2021)在关于宗族发展和个体主义/集体主义关系的研究中使用的统计指标，本研究选取了来自统计年鉴和人口普查数据的 2010 年的出生人口性别比和姓氏集中度（人口占比前 10 位的姓氏占总体的比例），以及量表测得的“宗族活动”指标，以地级市级别为空间粒度，将单个词的词频与以上三个指标进行相关分析。除 6 个地级行政单位因行政建制变化或统计信息缺失，问卷涵盖的 131 个地级市中有 125 个纳入分析。出于对时间因素和样本量的综合考虑，从微博样本中抽取了发布于 2011 年 1 月 1 日至 2012 年 12 月 31 日的微博文本数据用以相关分析。

根据宗族文化的基本内涵，假设高的出生人口性别比（男）、高的姓氏集中度和高的“宗族活动”得分能够正向预测本地区的宗族文化程度。

出于词表初步建立时对典型特征概括的考虑，相关分析中共选取了三个指标均处于前 27%的高宗族文化地区（11 个）和均处于后 27%的低宗族文化地区（9 个）。之后，选取了词频与三个指标中至少一个具有显著相关关系（$r > 0.1$，$p < 0.05$）的词，经筛选得到 273 个有效词语。

表 1 词表各部分及词的来源

词表	来源
外显标志词表 334	基于文献总结 106 个 亲属词 52 个 组合生成 144 个 另外添加的相关词 36 个
扩充词表 1316	由 32 个核心关键词人工扩充与大语言模型扩充得到
重复的词 70	重复出现在外显标志词表和扩充词表中

（3）词表的维度划分与筛选

利用 273 个有效词构建的词表，根据表 2 的宗族文化观念核心关键词，本研究确定了宗族文化观念体现在血缘认同、互助、孝道、传承、礼治、乡土观念六个具体方面及相关定义。

（4）词典效度验证

出于对词典使用的生态效度的考虑，研究中采用了 50%随机抽取和 50%基于关键词筛选的方式获取了 450 条微博文本，用以进行词频和人工评分对比检验。抽样中平衡了随机文本和关键词文本的内容长度。微博文本样本是从上述的 125 个地级市 2011、2012 年的数据总体中获取的。

围绕“宗族文化观念”和各维度的定义，建立了操作手册，要求评分者对“该文本是否表达了用户的宗族观念/宗族观念的具体维度”进行打分，采用 5 点里克特计分法，1 分表示非常不赞同，5 分表示非常赞同。评分手册中的定义和要求详见附表 2。之后首先抽取了 24 条文本以及 2 道认真度测试题进行一致性评价，5 位评分者（3 位为心理学硕士研究生，2 位为其他人士）都通过认真度测试，且评分达到了较高的一致性，kendall W 为 0.626（$p < 0.01$，$df = 167$）。

之后将 450 个文本分配给 5 位评分者，同时利用词频统计程序进行处理。经过人工评分和词频相关分析，结果发现词典覆盖率与人工宗族观念评分达到了正向显著相关，相关系数 0.231（$p < 0.01$）；血缘认同维度上达到正向显著的中等程度相关，相关系数 0.380（$p < 0.01$）；互助维度上达到正向显著的弱相关，相关系数 0.128（$p < 0.01$）；孝道维度上达到正向显著的弱相关，相关系数 0.219（$p < 0.01$）；传承维度上达到正向显著的中等程度相关，相关系数 0.369（$p < 0.01$）；礼治维度上达到正向显著弱相关，相关系数 0.091（$p = 0.027$）；乡土观念维度上达到正向显著的弱相关，相关系数 0.181（$p < 0.01$）。人工评分和词频结果相互之间及内部两两之间的相关关系

表 2 人工评分与词频统计结果皮尔逊相关矩阵

	人工-宗族观念	人工-血缘认同	人工-互助	人工-孝道	人工-传承	人工-礼治	人工-乡土观念
人工-宗族观念	--						
人工-血缘认同	.632**	--					
人工-互助	.414**	.409**	--				
人工-孝道	.465**	.486**	.263**	--			
人工-传承	.577**	.478**	.407**	.564**	--		
人工-礼治	.520**	.421**	.311**	.599**	.721**	--	
人工-乡土观念	.392**	.448**	.260**	.347**	.430**	.459**	--
	人工-宗族观念	人工-血缘认同	人工-互助	人工-孝道	人工-传承	人工-礼治	人工-乡土观念
词频-宗族观念	.231**	.162**	.189**	.157**	.300**	.172**	.124**
词频-血缘认同	.461**	.380**	.171**	.303**	.229**	.226**	.150**
词频-互助	0.069	0.057	.128**	-0.043	-0.022	-0.059	0.006
词频-孝道	.363**	.265**	.167**	.219**	.178**	.173**	0.069
词频-传承	.170**	0.061	.146**	.193**	.369**	.229**	.132**
词频-礼治	.079*	-0.01	0.008	0.045	.111**	.091*	0.021
词频-乡土观念	.249**	.186**	.116**	.176**	.189**	.151**	.181**
	词频-宗族观念	词频-血缘认同	词频-互助	词频-孝道	词频-传承	词频-礼治	词频-乡土观念
词频-宗族观念	--						
词频-血缘认同	.532**	--					
词频-互助	.627**	.261**	--				
词频-孝道	.456**	.788**	.294**	--			
词频-传承	.795**	.397**	.218**	.383**	--		
词频-礼治	.281**	.180**	0.031	0.045	.338**	--	
词频-乡土观念	.231**	.374**	0.069	.235**	.178**	.255**	--

2.4 结果

运用宗族观念词典，对全国 31 个省级行政单位在 2010 到 2022 年间的宗族观念词水平进行了统计。表 3 展示了宗族观念词水平在各年的最大值、最小值、算术平均值和中位数。图 1 显示了宗族观念词总体水平随时间变化的大体情况，总体上保持稳定，在 2013 年达到最高值，2014 到 2022 年有小幅的上升趋势。

显示“互助”维度词频在 2013 年达到最高，其后稍有下降；“传承”维度词频在 2010-2022 年间总体上升，在 2021 年达到最高；“血缘认同”“礼治”“孝道”均呈略微上升趋势，在 2022 年达到最高；“乡土观念”在研究设计的时间范围内无明显波动。

表 3　**2010-2022 年的全国宗族文化观念词水平**

年份	N	最大值	最小值	中位数	平均值
2010	31	0.0019	0.0010	0.0012	0.0012
2011	31	0.0014	0.0010	0.0012	0.0012
2012	31	0.0017	0.0011	0.0014	0.0014
2013	31	0.0029	0.0014	0.0017	0.0018
2014	31	0.0023	0.0011	0.0013	0.0014
2015	31	0.0020	0.0013	0.0015	0.0015
2016	31	0.0027	0.0012	0.0015	0.0015
2017	31	0.0030	0.0013	0.0016	0.0017
2018	31	0.0028	0.0011	0.0014	0.0015
2019	31	0.0025	0.0012	0.0016	0.0016
2020	31	0.0024	0.0010	0.0015	0.0016
2021	31	0.0028	0.0013	0.0016	0.0017
2022	31	0.0024	0.0013	0.0016	0.0017

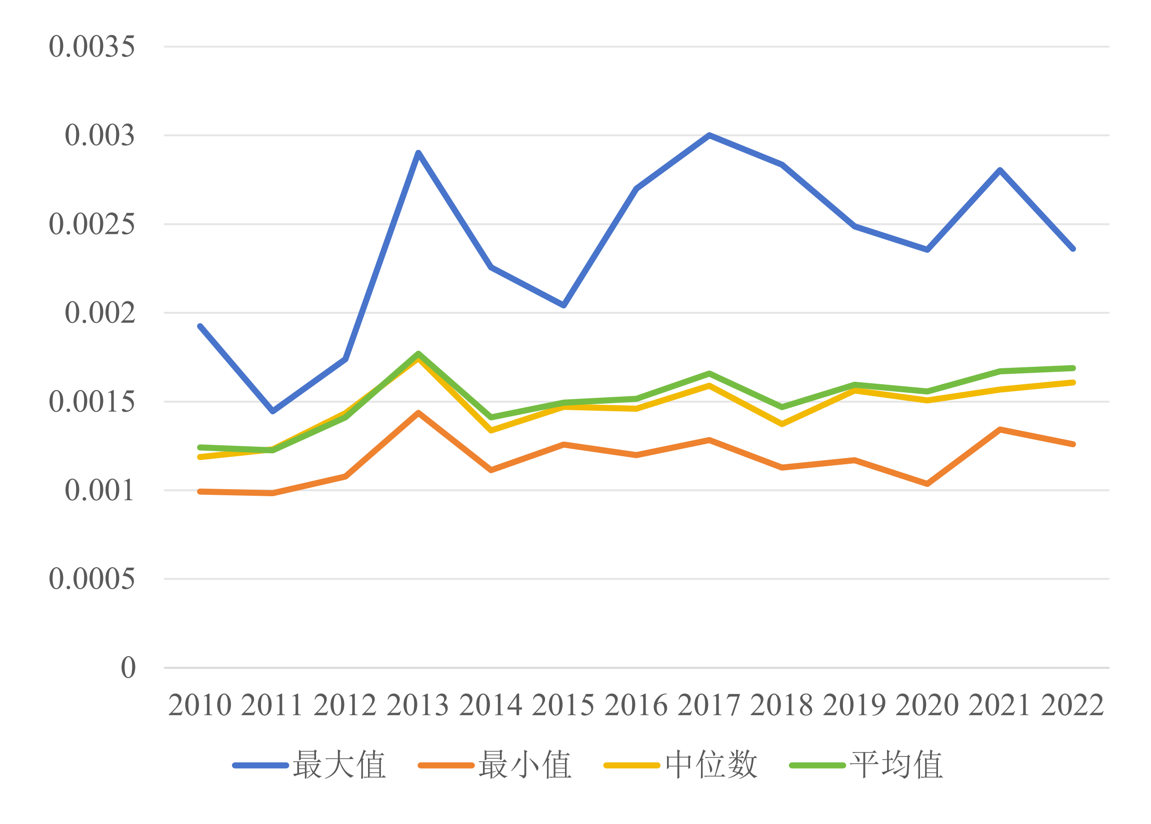


图 1　2010-2022 年的全国宗族文化观念词水平变化情况

同时还对各维度词频在各年的平均值、中位数作统计，见表 4。变化趋势如图 2，显示“互助”维度词频在 2013 年达到最高，其后稍有下降；“传承”维度词频在 2010-2022 年间总体上升，在 2021 年达到最高；“血缘认同”“礼治”“孝道”均呈略微上升趋势，在 2022 年达到最高；“乡土观念”在研究设计的时间范围内无明显波动。

表 4　2010-2022 年的全国宗族文化观念不同维度词频（单位：10^{-5}）

年份	N	血缘认同		互助		孝道	
		平均值	中位数	平均值	中位数	平均值	中位数
2010	31	8.50	8.47	38.55	37.32	4.71	4.58
2011	31	7.83	7.87	35.98	36.25	4.15	4.16
2012	31	7.83	7.70	45.83	44.54	3.98	3.94
2013	31	9.96	10.08	60.43	58.75	4.34	4.34
2014	31	8.18	8.24	41.96	38.37	4.13	4.12
2015	31	8.18	7.83	45.48	43.57	3.97	3.91
2016	31	8.02	7.70	43.05	42.27	4.02	3.93
2017	31	8.32	7.87	45.71	45.34	4.03	3.99
2018	31	8.92	8.49	35.80	36.14	3.99	3.92
2019	31	9.65	9.23	38.14	38.01	4.49	4.36
2020	31	9.19	8.99	38.98	37.41	4.27	3.96
2021	31	9.86	9.79	38.58	38.22	5.03	4.90
2022	31	21.81	21.40	36.99	34.51	17.02	16.68
年份	N	传承		礼治		乡土观念	
		平均值	中位数	平均值	中位数	平均值	中位数
2010	31	20.70	19.76	9.39	9.31	1.14	0.98
2011	31	22.70	22.34	10.52	10.40	1.35	1.29
2012	31	21.97	21.83	9.50	9.64	1.45	1.39
2013	31	25.25	25.24	11.96	11.53	1.34	1.28
2014	31	25.90	24.31	11.25	10.52	1.40	1.29
2015	31	25.95	24.37	11.83	11.11	1.38	1.25
2016	31	29.59	26.29	13.41	11.76	1.51	1.39
2017	31	34.43	31.11	14.65	13.51	1.43	1.30
2018	31	36.89	33.63	15.35	14.06	1.45	1.16
2019	31	46.30	43.57	17.67	16.00	1.23	1.09
2020	31	44.47	42.58	16.66	16.53	1.23	1.16
2021	31	50.58	46.40	19.78	17.05	1.22	1.10
2022	31	38.75	35.41	17.15	16.84	1.45	1.30

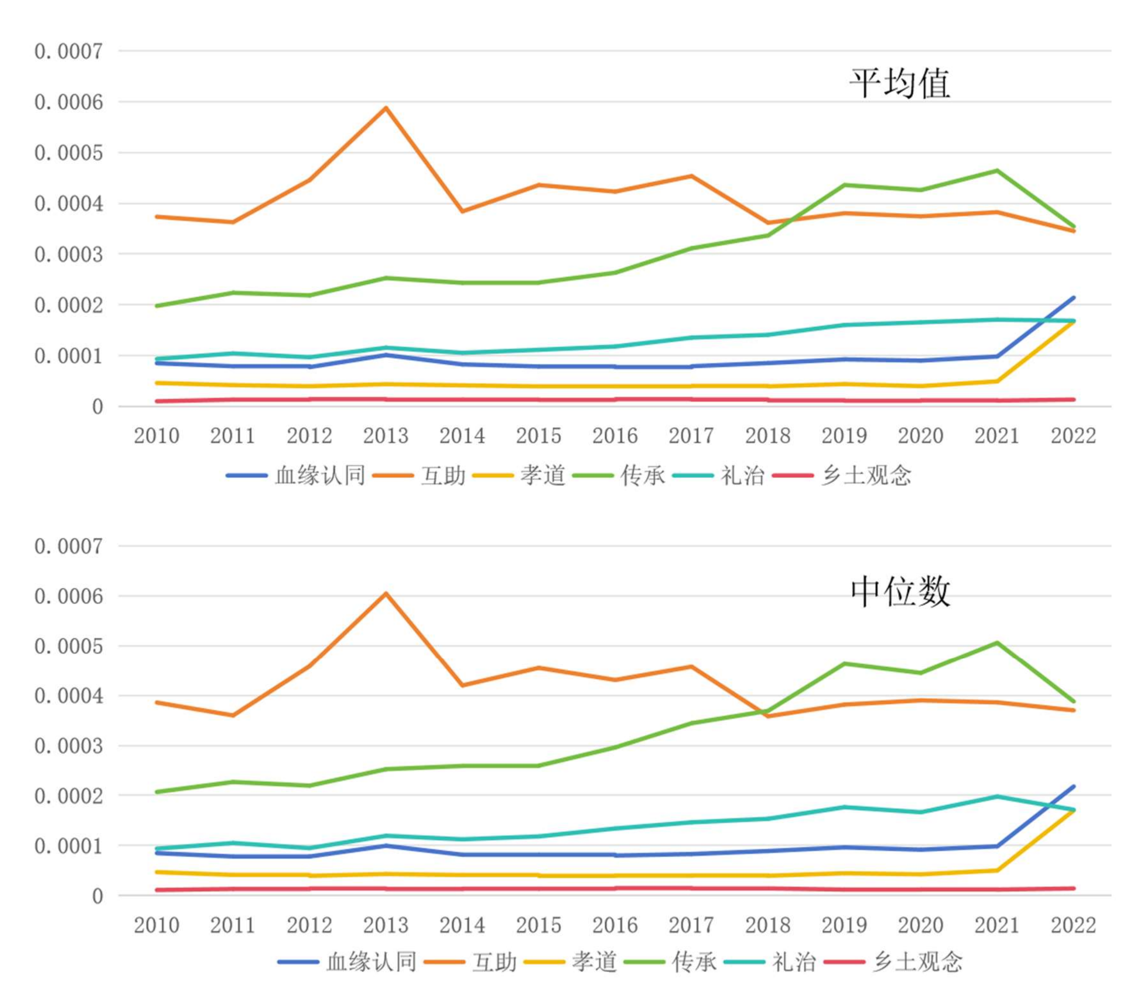


图 2 2010-2022 年的全国宗族文化观念不同维度词频变化趋势

以上结果显示 H1 得到了检验，基于词典的宗族观念测量方法可行，且与人类专家对于文本的评判具有显著相关性。基于 2010-2022 年的数据汇总显示，宗族文化观念水平在全国各省间有一定横向差异，而在时间纵向变化上则不明显。

3 研究二：宗族观念对家庭户小型化的作用

3.1 数据来源

本研究主要使用了利用 python 爬虫程序从新浪微博获取的文本数据，按照研究 1 获取的词典方法进行了汇总和词频分析。出于对数据规模和时间对其的考虑，使用了 2011、2012 年的中国 31 个省级单位（包含直辖市、自治区，不含港澳台）以地级市汇总的文本。并使用了中国国家统计局发布的中国统计年鉴(2011)和第六次人口普查数据(2010)。

3.2 变量意义

宗族观念：由于家庭户模式和结构是宗族文化观念相对总体的反映，总人口：地级市总人口代表着城市规模、集聚程度和现代化程度，人口标志着大型城市，即有着更现代化的家庭和居住观念、有更丰富的文化交流和更剧烈的人口流动。人均 GDP：代表着经济发展情况，被认为是家庭小型化的重要预测因素。非汉族人口比例：宗族文化观念被定义为与汉族的典型文化传统，因此对少数民族比例进行控制，避免族群异质性带来影响。非农人口比例：

中国户籍制度将一般家庭户区分为农业和非农业两种类型，而非农业人口比例通常作为衡量城市化水平的比例指标。经度、纬度：代表了不同的自然地理条件和生产模式，以及该地级市在中国地理分区中的基本位置。

交互项：各交互项意味着控制不同水平的某一经济地理因素情况下，宗族观念发挥的影响作用的差异。

3.3 结果

数据获取和处理后，利用 jamovi 软件包进行相关分析、多层广义回归模型分析。首先按照以单人家庭户占家庭户比例为因变量，纳入地级市总人口、人均 GDP、非汉族人口比例、非农人口占比、经度、纬度和宗族观念为预测变量，并考察了宗族观念与其他预测变量的交互项。

首先建立了基于家庭现代化理论的基础模型，考察了代表了经济发展水平和地理条件的自变量（地级市总人口、人均 GDP、非汉族人口比例、非农人口占比、经度、纬度）对单人家庭户比例的作用（模型 1）；之后以模型 1 为基准，每次加入一个新的预测变量，考察新的模型预测能力，分别是宗族观念（模型 2）、（模型 3）和宗族观念*地级市总人口（模型 4）。经过比较，此时再加入其它交互项的模型不再有显著提升。随后对模型 4 中回归效应显著的交互项进行简单效应分析。

多层回归比较结果见表 5，4 个模型的拟合参数情况见表 6。结果发现单独纳入宗族观念水平作为预测变量时无显著提升（$F = 1.03$，$p = 0.311$），而纳入宗族观念*人均 GDP 时（$F = 5.57$，$p = 0.019$）和纳入宗族观念*地级市总人口时（$F = 11.74$，$p < 0.001$）模型 R^2 有显著提升。此结果显示单独以宗族观念作为新的预测变量时并不能更好提高模型解释力，而同时考虑其与关键经济地理变量（人均 GDP、地级市总人口）的交互作用时能够显著增加解释力。

表 5 多层回归模型比较

比较							
模型		模型	*ΔR²*	*F*	*df1*	*df2*	*p*
1	-	2	0.00307	1.03	1	214	0.311
2	-	3	0.01624	5.57	1	213	0.019
3	-	4	0.03257	11.74	1	212	<.001

表 6 多层回归模型拟合总结

					模型总体检验			
模型	*R*	*R²*	*AIC*	*BIC*	*F*	*df1*	*df2*	*p*
1	0.6	0.36	-809	-782	20.2	6	215	<.001
2	0.603	0.363	-808	-778	17.4	7	214	<.001
3	0.616	0.379	-812	-778	16.3	8	213	<.001
4	0.642	0.412	-822	-785	16.5	9	212	<.001

将模型 4 作为有效容纳最多预测变量的固定效应预测模型（见表 7），发现对单人户比例起正向预测作用的经济地理变量有人均 GDP（$\beta = 0.20$，$p = 0.003$）、经度（$\beta = 0.29$，$p < 0.001$）、非农人口占比（$\beta = 0.18$，$p = 0.012$），起显著负向预测作用的经济地理变量有纬度（$\beta = -0.56$，$p < 0.001$）。宗族观念水平与单人户比例有着负向关联（$\beta = -0.16$，$p = 0.011$），而宗族观念∗人均 GDP（$\beta = -0.13$，$p = 0.019$）和宗族观念∗总人口（$\beta = -0.30$，$p < 0.001$）也有着显著的回归效应。通过对假设 H2 的检验，发现以本研究方法测得的宗族观念本身不能作为独立的预测变量，但在包含交互项的模型中有着显著的负向预测效应。

表 7 模型 4 固定效应参数估计

变量	系数	*SE*	*β*	*df*	*t*	*p*
(Intercept)	1.3E-01	2.5E-03	0.00	212	53.29	＜.001
总人口	3.1E-10	1.2E-09	0.02	212	0.26	0.792
人均 GDP	4.5E-07	1.5E-07	0.20	212	3.03	0.003
经度	1.1E-03	2.9E-04	0.29	212	3.79	＜.001
纬度	-3.5E-03	5.2E-04	-0.56	212	-6.77	＜.001
非农人口占比	6.0E-04	2.4E-04	0.18	212	2.53	0.012
非汉族占比	-1.2E-04	2.1E-04	-0.03	212	-0.59	0.557
宗族观念	-1.9E+01	7.2E+00	-0.16	212	-2.58	0.011
宗族观念*人均 GDP	-7.3E-04	3.1E-04	-0.13	212	-2.37	0.019
宗族观念*总人口	-1.3E-05	3.9E-06	-0.30	212	-3.43	＜.001

随后进行简单效应分析，以考察不同宗族观念水平下，人均 GDP 对单人户比例的作用差异。在低宗族观念水平（Mean-1SD）下，人均 GDP 能够正向预测单人户比例（$\beta=0.33$，$p < 0.001$），而在高宗族观念水平（Mean+1SD）下，人均 GDP 则没有显著效应（$\beta=0.06$，$p=0.49$），详见表 8 及图 3。

表 8 宗族观念调节下的地级市总人口简单效应估计

调节变量 宗族观念水平	系数	*SE*	*β*	*df*	*t*	*p*
Mean-1-SD	5.92E-09	1.67E-09	0.32	212	3.54	＜.001
Mean	3.05E-10	1.15E-09	0.02	212	0.26	0.79
Mean+1-SD	-5.31E-09	2.29E-09	-0.28	212	-2.32	0.02

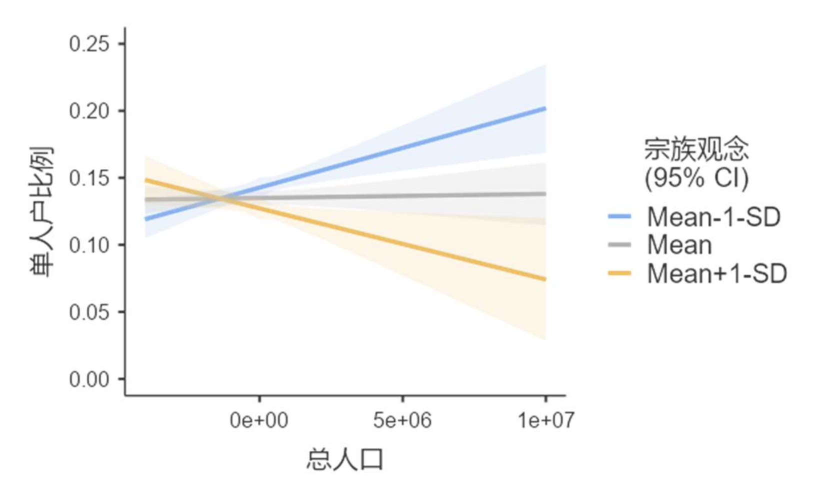


图 3 不同宗族观念水平下总人口对单人户比例的预测作用

对地级市总人口的简单效应分析显示，结果显示在低宗族观念水平（Mean-1SD）下，总人口能够正向预测单人户比例（$\beta=0.32$，$p < 0.001$），而在高宗族观念水平（Mean+1SD）下，总人口则是负向预测单人户比例（$\beta=-0.28$，$p=0.021$），详见表 9 及图 4。

表 9 宗族观念调节下的人均 GDP 简单效应估计

调节变量 宗族观念	Estimate	SE	β	df	t	p
Mean-1-SD	7.52E-07	1.86E-07	0.33	212	4.05	< .001
Mean	4.47E-07	1.48E-07	0.20	212	3.03	0.00
Mean+1-SD	1.42E-07	2.05E-07	0.06	212	0.69	0.49

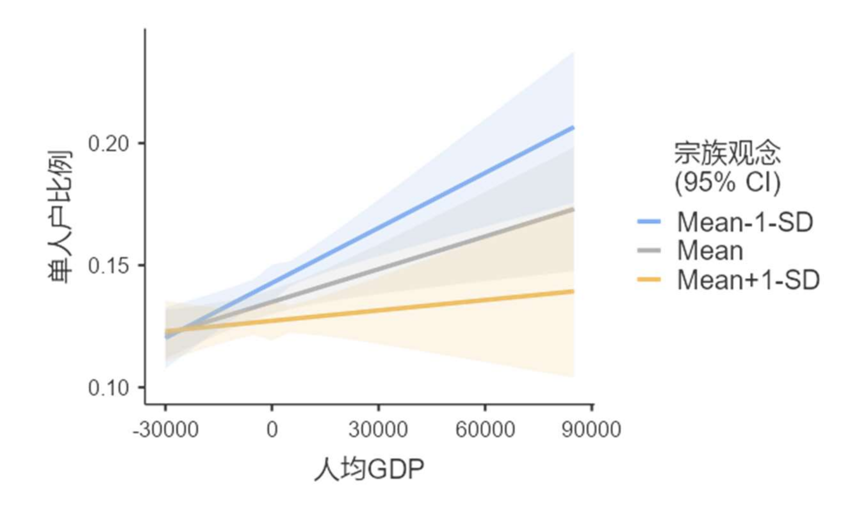


图 4 不同宗族观念水平下人均 GDP 对单人户比例的预测作用

通过以上分析，假设 H3 得到了检验，发现宗族观念确实对经济地理变量与单人户比例的关系起到调节作用。针对地级市总人口、人均 GDP 而言，都是随着宗族观念水平的变化，预测作用朝相反或减弱的方向变化。

4 结论与讨论

本研究通过基于历史文献的资料收集和基于访谈的概念形成方法，在宗族文化观念核心词的基础上进行了扩充、检验、筛选和维度划分，建立了包含六个维度的宗族文化观念词典，并经检验发现其具备满足推广应用的信效度。利用构建的词典对 2010 年至 2022 年间微博数据进行分析，发现宗族文化观念在全国范围内总体平稳。这表明在经历了快速城市化和现代化进程后，宗族文化观念仍在一定程度上影响着人们的价值观和行为。

研究还发现宗族文化观念对家庭户小型化产生了一定影响。具体来说，在建立了回归模型后，成功应用了通过微博测量的宗族文化观念水平作为预测变量。回归效应分析显示，宗族文化观念较强的地区，单人家庭户的比例较低。此外，经济发展水平、城市化进程和人口流动等因素与宗族文化观念的交互作用复杂，共同影响着单人户的地域分布模式。一方面我们的结果呼应了过去对单人户比例的家庭现代化解释，在变量作用上，人均 GDP 和非农人口占比较高的地区单人户比例较高，符合单人户作为更加现代化家庭模式的假设。此外，中国在经济发展的地理分布上，具有总体上西高东低、南高北低的特点，而回归分析结果中经度起到正向预测作用、纬度起到负向预测作用，与这一普遍认识保持一致。

这些结果说明关于单人家庭户的现代化假说仍是坚实有效的。但同时本研究揭示了宗族观念对这些主要经济地理因素的调节作用。其表现为在低宗族观念水平下，地级市总人口和人均 GDP 都对单人户起到正向预测作用，而随着宗族观念提高，它们不再作为正向预测因

素。具体来说，高宗族观念水平下，总人口起负向预测作用，而人均 GDP 不再是显著预测因素。这提示经济发展水平的作用仍然是通过文化和价值观念起作用的，在具有牢固的宗族观念传统的地区，经济发展并不必然引起家庭模式“定向演化”，即出现经济发展水平与家庭发展水平的“错位”（马春华等，2011）。因此有必要进一步深入探索宗族文化观念的内涵和表现，并利用其更好地揭示中国家庭演化中的文化观念因素作用。

本研究出于探索词典方法应用于宗族文化观念测量的可行性的考虑，主要局限于数据汇总和横截面分析，这限制了我们在因果关系方面的推断能力。未来的研究所应考虑采用纵向数据，追踪同一地区或个体在不同时间点上的变化，以更准确地理解宗族文化观念对家庭户结构变迁的长期影响，并可以期望识别出哪些因素在不同时间段内对单人户比例产生了显著影响。此外，虽然构建了一个包含六个维度的宗族文化观念词典，但在实际分析中主要关注了整体的宗族文化观念水平，而没有充分利用词典的多个维度。因此需要后续工作进一步阐明宗族文化观念代表的心理结构和发掘更坚实的基准变量，得到更丰富准确的理解。

本研究的发现进一步验证了文化影响对单人户地域差异的关键作用，且基于用户生成内容和大数据方法建立了有效的测量方法，开辟了一条宗族文化观念进行定量化实证分析的路径。未来可将词典及相应方法进一步用于家庭户结构和变化规律方面的研究，并在更小时空粒度上进行验证，以更好地回答单人户形成和变化规律的机制问题。

参考文献

[1] Gong, W., Zhu, M., Gürel, B., & Xie, T. (2021). The Lineage Theory of the Regional Variation of Individualism/Collectivism in China. Frontiers in Psychology, 11, 596762. https://doi.org/10.3389/fpsyg.2020.596762

[2] Peng, Y. (2004). Kinship Networks and Entrepreneurs in China’s Transitional Economy. American Journal of Sociology, 109(5), 1045–1074. https://doi.org/10.1086/382347

[3] Qiu, L., Lin, H., Ramsay, J., & Yang, F. (2012). You are what you tweet: Personality expression and perception on Twitter. Journal of Research in Personality, 46(6), 710–718. https://doi.org/10.1016/j.jrp.2012.08.008

[4] Cheung, A. K.-L., & Yeung, W.-J. J. (2015). Temporal-spatial patterns of one-person households in China, 1982–2005. Demographic Research, 32, 1209–1238.

[5] Dommaraju, P. (2015). One-person households in India. Demographic Research, 32, 1239–1266.

[6] Lee, S., Yoon, D., & Koh, J. (2015). A Study on the Age Distribution Factors of One Person Household in Seoul using Multiple Regression Analysis. Spatial Information Research, 23(3), 11–21.

[7] Li, T., Fan, W., & Song, J. (2020). The household structure transition in China: 1982–2015. Demography, 57, 1369–1391.

[8] Li, T., Yang, Y. C., & Zhang, Y. (2018). Culture, economic development, social-network type, and mortality: Evidence from Chinese older adults. Social Science & Medicine, 204, 23–30.

[9] Ronald, R. (2017). The remarkable rise and particular context of younger one–person households in Seoul and Tokyo. City & Community, 16(1), 25–46.

[10] Talhelm, T., Zhang, X., Oishi, S., Shimin, C., Duan, D., Lan, X., & Kitayama, S. (2014). Large-scale psychological differences within China explained by rice versus wheat agriculture. Science, 344(6184), 603–608.

[11] Waara, N., & Stjernborg, V. (2010). Mobility of older women and men at transition from a two person household to a one person household.

[12] Wang, X. (2007). Explaining the differences in the clan residing pattern between southern and northern Han Chinese. Research on Financial and Economic Issues, 288(11), 20–30.

[13] Wang, Y. (2008). The theoretical analysis for the family structure transition: Based on the historical and empirical evidence in rural China. Journal of Social Science, 7, 90–103.

[14] Wang, Y. (2013). The analysis of Chinese family structure change: Based on the 2010 population census data. Social Science in China, 12, 60–77.

[15] Wang, Y. (2014). An Analysis of changes in the Chinese family structure between urban and rural areas: On the basis of the 2010 National Census Data. Social Sciences in China, 35(4), 100–116.

[16] Yeung, W.-J. J., & Cheung, A. K.-L. (2015). Living alone: One-person households in Asia. Demographic Research, 32, 1099–1112.

[17] Yu, Z. (2017). Macro effects on the household formation of China’s young adults–demographics, institutional factors, and regional differences. International Journal of Housing Policy, 17(4), 512–540.

[18] Zeng, Y., & Wang, Z. (2018). Dynamics of family households and elderly living arrangements

in China, 1990–2010. China Population and Development Studies, 2(2), 129–157.
[19] 周长洪. (2013). 中国家庭结构变化的几个特征及其思考——基于 "五普" 和 "六普" 数据的比较. 南京人口管理干部学院学报, 29(4), 3–8.
[20] 王广州 & 周玉娇. (2021). 中国家庭规模的变动趋势, 影响因素及社会内涵. 青年探索, 4, 41–49.
[21] 王晓慧, 肖鹏峰, 康国定, & 杨萌萌. (2012). 中国分代家庭户的空间分布特征研究. 南方人口, 27(4), 61–72.
[22] 王跃生. (2016). 中国当代家庭,家户和家的"分"与"合". 中国社会科学, 000(004), 91–110.
[23] 王跃生. (2023). 当代单人户状态,变动及其特征分析. 人口与经济, 6, 38–55.
[24] 王跃生. (2024). 当代家庭户代际结构地区比较分析——以第七次人口普查数据为基础. 社会科学研究, 4.
[25] 胡湛 & 彭希哲. (2014). 中国当代家庭户变动的趋势分析——基于人口普查数据的考察. 社会学研究, 3, 145–166.
[26] 龚为纲. (2012). 中国农村分家模式的历史变动——基于 1990, 2000 年全国人口普查原始数据的分析. 青年研究, 4, 82–93.
[27] 吴祖鲲, 王慧姝. (2014). 文化视域下宗族社会功能的反思. 中国人民大学学报, 28(3), 132–139.
[28] 唐宗力. (2017). 城镇化与农村宗族文化. 安徽大学学报(哲学社会科学版), 41(3), 103–118. https://doi.org/10.13796/j.cnki.1001-5019.2017.03.014
[29] 姚伟钧. (2002). 宗法制度的兴亡及其对中国社会的影响. 华中师范大学学报(人文社会科学版), 3, 87–92.
[30] 常建华. (1999). 二十世纪的中国宗族研究. 历史研究, 5, 140–162.
[31] 李丹丹, 徐子超. (2022). 宗族文化对人口流动范围的影响——基于族谱、祠堂与 CMDS 数据分析. 南方人口, 37(2), 48–59.
[32] 王玉珊. (2018). 澡塘村宗族观念在丧葬习俗中的传承研究 [硕士, 云南民族大学].
[33] 王询. (2007). 中国南北方汉族聚居区 宗族聚居差异的原因. 财经问题研究, 11, 20–30.
[34] 苏悦, 刘明明, 赵楠, 刘晓倩, 朱廷劭. (2021). 基于社交媒体数据的心理指标识别建模:机器学习的方法. 心理科学进展, 29(4), 571–585.
[35] 陈丽. (2007). 转型时期农村宗族势力复兴研究 [硕士, 华中师范大学].
[36] 徐勇, & 邓大才. (2017). 中国农村调查(总第 2 卷•村庄类第 1 卷•华南区域第 1 卷): 中国农村调查. 总第 9 卷. 村庄类第 8 卷. 长江区域第 1 卷.
[37] 李泓仪，董一萱，申张辰然，俞佳妍.（未发表）宗族文化观念量表编制.
[38] 纪馨玉,董一萱,朱廷劭.(2023).中国传统宗族文化的表达模式与变化.中国科学院科技论文预发布平台.doi:10.12074/202311.00095V2
[39] 中国国家统计局. (2010). 中国统计年鉴・2010. 中国统计出版社.
[40] 中国国务院第六次全国人口普查办公室. (2011). 2010 年第六次全国人口普查主要数据. 中国统计出版社.

(通讯作者: 朱廷劭 Email:tszhu@psych.ac.cn)

Exploring the Relationship Between Clan Cultural Concepts and the Geographical Differences in Single-Person Households Based on Weibo Data

ZHOU Zili[1,2], LI Hongyi[1,2], ZHU Tingshao[1,2]
ˆ1 (Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China)
ˆ2 (Department of Psychology, University of Chinese Academy of Sciences, Beijing 100049, China)

[Abstract] In recent years, the proportion of single-person households has risen rapidly, reflecting the ongoing changes in family and living patterns in China, which are likely to be profoundly influenced by cultural factors, particularly traditional Chinese clan culture. This study attempts to introduce a dictionary-based statistical method using Weibo text data, combined with principles from psycholinguistics, to measure the level of clan cultural concepts expressed in user-generated content. It also explores the role of clan cultural concept levels in the geographical differences among single-person households by integrating other variables. The results of the clan concept dictionary measurement showed a considerable correlation with manual ratings, meeting the initial requirements for analyzing Weibo text data. In cross-sectional regression analysis, it was found that the regression model significantly improved when clan cultural concepts and their interactions with city-level population size and per capita GDP were included simultaneously. Moreover, there is a negative association between the level of clan cultural concepts and the proportion of single-person households. The findings of this study suggest the role of clan culture in the formation of geographical differences in single-person households, providing new insights into bridging the theoretical and empirical gaps in the transformation of family patterns. It also paves the way for the quantitative use of clan cultural indicators in future research.
[Keywords] Clan Culture, Single-Person Households, Social Media